

PCT National Publication Gazette

National Patent Publication No. 11-513144  
Date of National Publication: November 9, 1999  
International Class(es): G09B 7/04  
5/06  
19/04  
19/06

(34 pages in all)

---

Title of the Invention: Apparatus for Interactive Language Training  
Patent Appln. No. 10-505803  
Filing Date: May 4, 1997  
Date of Filing Translation: March 11, 1998  
International Filing No. PCT/IL97/00143  
International Publication No. WO98/02862  
International Publication Date: January 22, 1998  
Priority Claimed: Country: U.S.A.  
Filing Date: July 11, 1996  
Serial No. 08/678,229  
Inventor(s): SHPIRO, Zeev  
Applicant(s): DIGISPEECH (ISRAEL) LTD.

(transliterated, therefore the spelling might be incorrect)

BEST AVAILABLE COPY

(19)日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11)特許出願公表番号

特表平11-513144

(43)公表日 平成11年(1999)11月9日

(51)Int.Cl.  
G 0 9 B 7/04  
5/06  
19/04  
19/06

識別記号

F I  
G 0 9 B 7/04  
5/06  
19/04  
19/06

審査請求 未請求 予備審査請求 未請求(全 34 頁)

(21)出願番号 特願平10-505803  
(86) (22)出願日 平成9年(1997)5月4日  
(85)翻訳文提出日 平成10年(1998)3月11日  
(86)国際出願番号 PCT/IL97/00143  
(87)国際公開番号 WO98/02862  
(87)国際公開日 平成10年(1998)1月22日  
(31)優先権主張番号 08/678, 229  
(32)優先日 1996年7月11日  
(33)優先権主張国 米国 (U S)

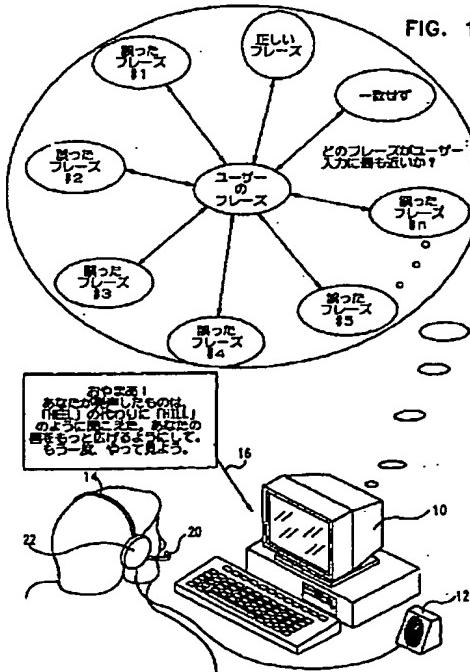
(71)出願人 ディジスピーチ・(イスラエル)・リミテッド  
イスラエル国テル・アヴィヴ 69513, ケヒラット・サロニキ・ストリート 13  
(72)発明者 シビロ, ズィーヴ  
イスラエル国テル・アヴィヴ 69122, ベイト・ズリ・ストリート 8  
(74)代理人 弁理士 社本 一夫 (外5名)

最終頁に続く

(54)【発明の名称】 対話型言語トレーニング装置

(57)【要約】

本発明は、ユーザによる予期される音響的応答を引出すトリガー発生器と、満足し得る発音を含む第1の複数の予期応答基準と、満足し得る発音を含む第1の複数の予期応答基準の各々に対する各々が異なる発音の誤りを含む第2の複数の予期応答基準とを含む、複数の予期応答基準を含む予期される音響的応答基準ライブラリと、ユーザにより行われる予期される音響的応答と基準が予期する応答との間の関係を示す音響的応答スコアリングと、ユーザによる行われた予期される音響的応答における発音の誤りをユーザに対して示すユーザ・フィードバック・インターフェース (12, 14, 16) を含む対話型言語教育装置である。本発明はまた、少なくとも第1および第2の言語の発声要素を含む少なくとも1つのデータベースと、認識される話された発声を受取るレシーバと、前記話された発声の特徴を少なくとも第1および第2の言語の前記発声要素の特徴の組合せと比較するコンパレータとを含む音声認識装置を開示する。ある場合に、発声要素の組合せが単一の発声要素を含むことが判る。音声認識の方法もまた開示される。



**【特許請求の範囲】**

1. ユーザによる予期される音響的応答を引出すトリガー発生器と、  
満足し得る発音を有する第1の複数の予期される応答基準と、該満足し得る発音を有する第1の複数の予期される応答基準の各々に対してそれぞれ異なる発音の誤りを含む第2の複数の予期される応答基準とを含む、複数の予期される応答基準を含む予期される音響的応答基準ライブラリと、  
ユーザが与える予期される音響的応答と予期される応答基準との間の関係を示す音響的応答スコアラと、  
ユーザが与える予期される音響的応答における発音の誤りをユーザに対して示すユーザ・フィードバック・インターフェースと、  
を備える対話型言語訓練装置。
2. 前記ユーザ・フィードバック・インターフェースが、発音の誤りを克服する方法についてのユーザに対する教育をも行う請求項1記載の装置。
3. 前記ユーザ・フィードバック・インターフェースが、予期される各音響的応答直後の各発音の誤りをユーザに対して示す請求項1記載の装置。
4. 前記ユーザ・フィードバック・インターフェースが、前記発音の誤りの音響的および視覚的な表示を行う請求項1記載の装置。
5. 予期される音響的応答が音響的手本の反復となるように音響的手本発生器が動作する請求項1記載の装置。
6. 予期される音響的応答が前記音響的手本の反復以外となるように前記音響的手本発生器が動作する請求項1記載の装置。
7. 予期される音響的応答が1つ以上のあり得る予期される音響的応答から選択される音響的手本となるように前記音響的手本発生器が動作する請求項1記載の装置。
8. 前記トリガー発生器が音響的手本をユーザに対して再生する音響的手本発生器を含む請求項1記載の装置。
9. 前記トリガー発生器が、視覚的トリガー出力をユーザに対して与える視覚的トリガー発生器を含む請求項1記載の装置。

10. 前記予期される音響的ライブラリが、予期される音響的応答基準データベースを含む請求項1記載の装置。
11. 前記予期される音響的応答基準データベースが、複数のテンプレートを含む請求項10記載の装置。
12. 前記予期される音響的応答基準データベースが、話者から独立的である請求項10記載の装置。
13. 前記予期される音響的応答基準データベースが、話者から独立的である請求項11記載の装置。
14. ユーザによる予期される音響的応答を引出すステップと、  
満足し得る発音を含む第1の複数の予期される応答基準と、該第1の複数の予期される応答基準の各々に対して各々が異なる発音の誤りを含む第2の複数の予期される応答基準と、を含む複数の予期される応答基準を含む予期される音響的応答基準ライブラリを提供するステップと、  
ユーザが与えた予期される音響的応答と基準が予期する応答との間の関係を示すステップと、  
ユーザが行う予期される音響的応答における発音の誤りをユーザに対して示すステップと  
を含む対話型言語訓練方法。
15. 発音の誤りを克服する方法をユーザに対して教示するステップをも含む請求項14記載の方法。
16. 各々の予期される音響的応答直後に各発音の誤りをユーザに対して示すステップをも含む請求項14記載の方法。
17. 前記発音の誤りの音響的および視覚的な表示を前記ユーザに対して与えるステップをも含む請求項14記載の方法。
18. 前記予期される音響的応答が前記音響的手本の反復である請求項14記載の方法。
19. 前記予期される音響的応答が前記音響的手本の反復以外である請求項14記載の方法。

20. 前記予期される音響的応答が、1つ以上のあり得る予期される音響的応答から選択される音響的手本である請求項14記載の方法。

21. 音響的応答を引出す前記ステップが、音響的手本をユーザに対して再生することを含む請求項14記載の方法。

22. 音響的応答を引出す前記ステップが、視覚的トリガー出力をユーザに対して与えることを含む請求項14記載の方法。

23. 少なくとも第1および第2の言語の発声要素を含む少なくとも1つのデータベースと、  
認識されるべき話された発声を受取るレシーバと、  
前記話された発声の特徴と少なくとも第1および第2の言語の前記発声要素の特徴の組合せと比較するコンパレータと、  
を備える音声認識装置。

24. ユーザによる予期される音響的応答を引出すトリガー発生器と、  
少なくとも第1および第2の言語の発声要素を含む少なくとも1つのデータベースと、認識される話された発声を受取るレシーバと、前記話された発声の特徴と少なくとも第1および第2の言語の前記発声要素の特徴の組合せと比較するコンパレータと、を含むユーザにより発声された予期される音響的応答を受取る発声認識器と、  
ユーザにより発声された予期される音響的応答における誤りをユーザに対して示すユーザ・フィードバック・インターフェースと、  
を備える言語教育システム。

25. 前記発声要素が、音素と、二音素と、音素間の遷移との少なくとも1つを含む請求項23記載の言語教育システム。

26. フレーズ・テンプレートを生成するよう動作するテンプレート発生器をも備える請求項23記載の言語教育システム。

27. 前記レシーバにより受取られる話された発声の特徴を抽出するよう動作する特徴抽出器をも備える請求項23記載の言語教育システム。

28. 少なくとも第1および第2の言語の発声要素を含む少なくとも1つのデータ

タベースを提供するステップと、

認識される話された発声を受取るステップと、

前記話された発声の特徴と少なくとも第1および第2の言語の前記発声要素の特徴の組合せと比較するステップと  
を含む音声認識方法。

29. 前記話された発声が第2の言語のネイティブ・スピーカであるユーザにより第1の言語で話され、前記少なくとも1つのデータベースが第1および第2の両方の言語の発声要素を含む請求項28記載の音声認識方法。

30. 前記少なくとも第1および第2の言語が異なる国語を含む請求項28記載の方法。

31. 前記少なくとも第1および第2の言語が单一の国語の異なる方言を含む請求項28記載の方法。

## 【発明の詳細な説明】

## 対話型言語トレーニング装置

## 発明の分野

本発明は、特に教育システムにおけるアプリケーションを備えた音声認識システムに関し、特に音素に基く音声認識（phoneme based speech recognition）を提供して言語を教えるためのコンピュータ化システムに関する。

## 発明の背景

言語教育用のコンピュータ化システムは公知である。一人が本発明の発明者である米国特許第5,487,671号には、とりわけユーザの言語の基準に対する関係の表示を行う言語教育のためのコンピュータ化システムが記載されている。

実質的に同じ特徴を有する製品が、商品名「英語を話せるように習おう（Learn to Speak English）」の下にThe Learning社から購入可能である。

この分野における他の市販製品が、HyperGlot Berlitz、Language Systems Mindscape Global LanguageおよびRosetta Stone Language Libraryから入手可能である。

音素に基く音声認識のためのコンピュータ化システムは公知であり、市販されている。かかるシステムの事例は、下記のものを含む。即ち、

IBM発売の「IBM VoiceType, 学生、家庭および小規模ビジネス用の簡単会話（Simply Speaking for students, home users and small business）」、

IBM発売の「IBM VoiceType, 専門用およびビジネス用（for professional and business use）」、

米国マサチューセッツ州NewtonのDragon Systems発売の「私に話しかけて（Talk To Me）」、

ベルギー L e p e r の L e r n o u t & H a u s p i e S p e e c h  
P r o d u c t s N. V. 発売の「A S R - 1 5 0 0」

発明の概要

本発明は、ユーザが犯す発音の誤りの種類のユーザに対する表示を行う言語教育用の更に改善されたコンピュータ化システムの提供を目的とする。

このように、本発明の装置の望ましい実施の形態によれば、  
ユーザの予期される音声応答を引出すトリガー発生器と、  
満足し得る発音を有する第1の複数の予期される応答基準と、該満足し得る発音を有する第1の複数の予期される応答基準の各々に対してそれが異なる発音の誤りを含む第2の複数の予期される応答基準とを含む、複数の予期される応答基準を含む予期される音響的応答基準ライブラリと、

ユーザが与える予期される音響的応答と予期される応答基準との間の関係を示す音声応答スコアと、

ユーザが与える予期される音声応答における発音の誤りをユーザに対して示すユーザ・フィードバック・インターフェースと、  
を備える対話型言語訓練用の装置が提供される。

ユーザ・フィードバック・インターフェースは、発音の誤りを克服する方法をユーザに対して指導をも行うことが望ましい。

本発明の望ましい実施の形態によれば、ユーザ・フィードバック・インターフェースが、それぞれ予期される音響的応答の直後に各発音の誤りをユーザに対して示す。

フィードバック・インターフェースは、発音の誤りの音響的および視覚的な表示を行うことが望ましい。

本発明の望ましい実施の形態によれば、予期される音響的応答が音響的手本の繰返しとなるように音響的手本発生器が動作する。

あるいはまた、音響的手本発生器は、予期される音響的応答が音響的手本の繰返しにとどまらないように動作する。

更なる代替例として、音響的手本発生器は、予期される音響的応答が1つ以上

のあり得る予期される音響的応答から選択される音響的手本であるように動作する。

トリガー発生器は、ユーザに対して音響的手本を再生するための音響的手本発生器を含むことが望ましい。

あるいは更に、トリガー発生器は、ユーザに対して視覚的トリガー出力を行うための視覚的トリガー発生器を含む。

予期される音響的応答ライブラリは、予期される音響的応答基準データベースを含むことが望ましい。

本発明の望ましい実施の形態によれば、予期される音響的応答基準データベースは、複数のテンプレートを含み、話者から独立的である。

また、本発明の望ましい実施の形態によれば、

ユーザによる予期される音響的応答を引出すことと、

満足し得る発音を有する第1の複数の予期される応答基準と、該満足し得る発音を有する第1の複数の予期される応答基準の各々に対してそれが異なる発音の誤りを有する第2の複数の予期される応答基準と、を含む複数の予期される応答基準を含む予期される音響的応答基準ライブラリを提供することと、

ユーザが与えた予期される音響的応答と基準が予期する応答との間の関係を表示することと、

ユーザが行った予期される音響的応答における発音の誤りをユーザに対して示すことと

を含む対話型言語訓練方法が提供される。

更に本発明の望ましい実施の形態によれば、この方法は、発音の誤りを克服する方法をユーザに対して指導を行うことをも含む。

更にまた本発明の望ましい実施の形態によれば、この方法は、各々の予期される音響的応答直後のそれぞれの発音の誤りをユーザに対して示すことをも含

む。

更に本発明の望ましい実施の形態によれば、この方法は、前記発音の誤りの音響的および視覚的な表示を前記ユーザに対して行うことをも含む。

更に本発明の望ましい実施の形態によれば、この方法は、予期される音響的応答が前記音響的手本の繰返しであることをも含む。

あるいはまた、この方法は、予期される音響的応答が前記音響的手本の繰返しにとどまらないことをも含む。

更に本発明の望ましい実施の形態によれば、予期される音響的応答が、1つ以上のあり得る予期される音響的応答から選択される音響的手本である。

更にまた本発明の望ましい実施の形態によれば、音響的応答を引出すステップが音響的手本をユーザに対して再生することを含む。

その上更に本発明の望ましい実施の形態によれば、音響的応答を引出すステップが視覚的トリガー出力をユーザに対して行うことを含む。

また、本発明の望ましい実施の形態によれば、少なくとも第1および第2の言語の音声要素を含む少なくとも1つのデータベースと、認識される発声された音声を受取るレシーバと、少なくとも第1および第2の言語の音声要素の特徴の組合せと発声された音声の特徴を比較するコンパレータとを含む音声認識装置もまた提供される。ある場合には、音声要素の特徴の組合せが1つの音声要素の特徴を含み得ることが判る。音声要素の特徴は、音声要素信号を含み得る。

また、本発明の望ましい実施の形態によれば、ユーザによる予期される音響的応答を引出すトリガー発生器と、および少なくとも第1および第2の言語の音声要素を含む少なくとも1つのデータベースと、認識される発声された音声を受取るレシーバと、少なくとも第1および第2の言語の前記音声要素の特徴の組合せと前記発声音声の特徴を比較するコンパレータとを含むユーザにより発声された予期される音響的応答を受取る発声認識器と、およびユーザにより発声された予期される音響的応答における誤りをユーザに対して示すユーザ・フィードバック・インターフェースとを含む言語教育システムもまた提供される。ある場合に、音声要素の特徴の組合せが1つの音声要素の特徴を含み得ることが判る。音声

要素の特徴は、音声要素信号を含み得る。

更に本発明の望ましい実施の形態によれば、音声要素は、音素、二音素および

音素間の遷移の少なくとも1つを含む。

更にまた本発明の望ましい実施の形態によれば、言語教育システムは、フレーズ・テンプレートを生成するよう動作するテンプレート発生器をも含む。

更に本発明の望ましい実施の形態によれば、言語教育システムは、レシーバにより受取られる発声された音声の特徴を抽出するよう動作する特徴抽出器とも含む。

また、本発明の望ましい実施の形態によれば、少なくとも第1および第2の言語の音声要素を含む少なくとも1つのデータベースを提供し、認識される発声音声を受取り、少なくとも第1および第2の言語の音声要素の特徴の組合せと発声音声の特徴を比較することを含む音声認識方法もまた提供される。ある場合に、音声要素の特徴の組合せが1つの音声要素の特徴を含むことが判る。音声要素の特徴は、音声要素信号を含む。

更にまた本発明の望ましい実施の形態によれば、発声音声は、第2の言語のネイティブ・スピーカであるユーザにより第1の言語で話され、この場合、少なくとも1つのデータベースが第1および第2の言語の両方の音声要素を含む。

更にまた本発明の望ましい実施の形態によれば、少なくとも第1および第2の言語は異なる国語を含む。

更に本発明の望ましい実施の形態によれば、少なくとも第1および第2の言語は1つの国語の異なる方言を含む。

#### 図面の簡単な説明

本発明は、図面に関して以降の詳細な記述を参照することにより更に完全に理解されよう。

図1は、本発明の望ましい実施の形態により構成され動作する対話型言語教育システムの全体的概略図、

図2は、言語の教育中の図1のシステムの動作の全体的機能ブロック図、

図3は、本発明の一実施形態による音響的基準ライブラリ生成中の図1のシステムの動作の全体的機能ブロック図、

図4は、本発明の別の実施の形態による音響的基準ライブラリ生成中の図1の

システムの動作の全体的機能ブロック図、

図5Aおよび図5Bは、図2の全体的機能ブロック図による言語教育中のシステムの動作を示す全体的フローチャートを構成し、

図6A、図6Bおよび図6Cは、図3の全体的機能ブロック図による言語教育のための音響的基準ライブラリ生成中のシステムの動作の1つの方法を示す全体的フローチャートを構成し、

図7は、図4の全体的機能ブロック図による言語教育のための音響的基準ライブラリ生成中のシステムの動作を示す全体的フローチャート、

図8は、図4に用いられた形式の音素テンプレート・データベースの生成の簡略表示、

図9は、ラベルを付した音声波形の簡略表示、

図10は、本発明の望ましい実施の形態による多言語の音素データベースの生成を示し、

図11は、音素を用いる音声認識の図、および

図12は、種々の言語の音素を用いる音声認識の図である。

#### 望ましい実施の形態の詳細な記述

まず、本発明の望ましい実施の形態により構成され動作する対話型言語教育システムの全体的概略図である図1と、図1のシステムの言語教育中の動作の全体的機能ブロック図である図2とを参照する。

図1のシステムが、開示が参考のため本文に援用される米国特許第5, 487, 671号に記載された発声教育用のコンピュータ化システムと多くの類似点を有することを理解すべきである。

以下に詳細に述べるように、本発明のシステムは、それぞれ異なる発音の誤りを持つ予期応答基準 (reference expected response) で動作し、かつユーザから与えられる予期音響的応答と発音の誤りを持つ該予期応答基準との間の関係を示す音響的応答スコアを含むことにおいて米国特

許第5, 487, 671号のシステムと異なる。

図1および図2のシステムは、本発明の望ましい実施の形態による音声認識機

能を組んでいる。

図1および図2のシステムは、33MHz以上で動作するIntel 8048 6CPUを用い、少なくとも8MHzのメモリを持ち、DOSバージョン6.0以上のオペレーティング・システムが走るIBM PCまたは互換機の如き周知のパーソナル・コンピュータに基くものであることが望ましい。パーソナル・コンピュータ10は、補助的音響モジュール12を備えることが望ましい。例えば、適切な音響モジュール12は、Digispeech社製であって米国カルフォルニア州Mountain ViewのDSP SOLUTIONS社により米国内で供給されるDigispeech Plusオーディオ・アダプタ(DS311)である。

一般に、パーソナル・コンピュータ10および音響モジュール12は、下記の機能を提供するように適切なソフトウェアが供給される。即ち、

予期されるユーザの音響的応答を引出すトリガー発生器。このトリガー発生器は、ユーザに対して音響的手本を再生するための音響的手本発生器を含むことが望ましいが、更にあるいは代替的に、ユーザに対して視覚的トリガー出力を提供するための視覚的トリガー発生器とを含み得る。

複数の予期応答基準(reference expected responses)を含み、該複数の予期応答基準が、満足し得る発音を有する第1の複数の予期応答基準と、該満足し得る発音を有する第1の予期応答基準の各々に対して各々が異なる発音エラーを有する第2の複数の予期応答基準であって、種々の言語の音素から構成される応答を含み得、そして音声認識における一般的の用途を有し得る該第2の複数の予期応答基準と、を含む予期される音響的応答基準ライブラリ。

ユーザが与える予期音響的応答と予期応答基準との間の関係を示す音響的処理スコアラ。

ユーザが与える予期音響的応答においてユーザに対して発音の誤りがあればそ

の誤りを示すユーザ・フィードバック・インターフェース。

ユーザ・フィードバック・インターフェースは、音響モジュール12とヘッド

セット14とを介して音響的フィードバックを提供することが望ましい。更に、図1および図2に示されるように、ユーザに対して視覚的に発音の誤りを示す、例えば図1に示される如きディスプレイ16が設けられることが望ましい。

本発明の望ましい実施の形態によれば、合計で6つの異なるデータベースが用いられる。本発明の便利かつ容易な理解のため、6つのデータベースについては、それらが本発明において生成され使用される順序で以下において簡単に述べる。

A. 暫定音響的手本データベース (Interim Audio Specimens Database) : このデータベースは、種々の地理的範囲、種々の年齢および両方の性の配分を含む複数のネイティブ・スピーカを記録することにより生成される。複数のネイティブ・スピーカは、種々の異なる言語を話す話者を含む。各話者は、複数の予め定めたフレーズを発音する。各話者は、複数の予め定めたフレーズのそれぞれに対して、フレーズを正しく発音し、またその都度複数の予め定めたフレーズのそれぞれに対して異なる誤りで数回、フレーズを誤って繰返す。このデータベースは、強化された統計的基礎を提供するように、各話者ごとに先に発音された各フレーズの複数の記録を含むことが望ましい。

B. 予期音響的応答基準データベース (Expected Audio Response Reference Database) : これは、記録された発声ではないテンプレート (template) を含むデータベースである。

色々な種類のテンプレートが提供される。単語ベースの音声認識において有効な1つの種類のテンプレートは、以下に述べる方法で「データベースA」から得られる。音素ベースの音声認識において有効な別の種類のテンプレートは、1つのフレーズと一緒に表わす発声要素の特徴の色々な組合せを含む。

単語ベースの音声認識において有効なテンプレートは、発音されたフレーズのそれぞれの発声パラメータを抽出して、先に述べた複数のネイティブ・スピーカの発音を表わすようにこれらパラメータを統計的に組合せることによって、暫

定音響的手本データベースAから得られる。

このように、各テンプレートは、ネイティブ・スピーカのグループの発音の統

計的組合せを表わす。

暫定音響的手本データベースAに発音が記録される全てのネイティブ・スピーカを網羅するように、1つのテンプレートを生成することができ、あるいは1つのテンプレートが全ての範囲のネイティブ・スピーカを正確に表わさない時は、複数のテンプレートを使用することもできる。例えば、1つのテンプレートが男性と他の女性とを表わすこともできる。代替的に、あるいは更に、個々のテンプレートがそれぞれ異なる言語の音素を含むこともできる。

本発明の望ましい実施の形態によれば、「予期音響的応答基準データベースB」が、先に述べた予期音響的応答の基準ライブラリを構成する。これは、話者から独立したデータベースである。

色々な種類のテンプレートを提供することができる。単語ベースの音声認識において有効な1つの種類のテンプレートは、先に述べた方法でデータベースAから得ることができる。音素ベースの音声認識において有効な別の種類のテンプレートは、1つのフレーズと一緒に表わす発声要素の特徴の色々な組合せを含むものである。

C. 音素データベース (Phonetic Database) : これは、所与の言語に対する音素の発声パラメータの市販されるデータベースである。このようなデータベースは、例えば、AT & T、米国コロラド州BoulderのSpeech Systems社、およびベルギーLeperのLernout & Hauspie Speech Products N. V. から入手可能である。それぞれ異なる言語の音素の発声パラメータを含む多重音素データベースが提供され、まとめて音素データベースと呼ばれている。

D. ユーザ・フォローアップ・データベース (User Followup Database) : これは、記録されたユーザ応答のコレクション (collection) である。

E. 予期音響的手本データベース (Expected Audio Speci-

men Database) : これは、複数のフォローアップの各々を正しく発音する一人の訓練話者の各々の記録のコレクションである。

F. 基準音響的手本データベース (Reference Audio Specimen Database) : これは、それぞれが複数の予め定めた発音誤りの異なる 1 つを含む複数のフレーズの各々を数回不正確に発音する一人の訓練した話者の記録のコレクションである。

次に、言語の教育中に図 1 のシステムの動作の全体的機能ブロック図である図 2 を参照する。

ユーザによる予期音響的応答を引出すため、「予期音響的手本データベース E」に格納された音響的手本が、音響モジュール 14 (図 1) を介してユーザに対して再生される。「ユーザ・フォローアップ・データベース D」に格納されるユーザの音響的応答を記録するため、通常はヘッドセット 14 の一部であるマイクロフォン 20 が用いられる。音響的手本は、典型的に発声されたフレーズを含む。これらフレーズは、1 つ以上の単語を含む。代替的に、あるいは更に、ユーザから予期音響的応答を引出すためユーザへ視覚的トリガー出力を提供するための視覚的トリガー発生器が設けられる。

発声されたフレーズ・パラメータがユーザの音響的応答から抽出され、ユーザの音響的応答の発声されたフレーズ・パラメータと「予期音響的応答基準データベース B」に格納された対応する正確あるいは不正確なフレーズの基準フレーズ・パラメータとの間の一致の類似性を計測するため、基準フレーズ・パラメータと比較される。

基準フレーズ・パラメータが必ずしも単語または単語の組合せを含む必要がないことが判る。その代わり、基準フレーズ・パラメータは、特に音素ベースの音声認識が実施されている時に、発声要素の特徴の種々の組合せを含むことができる。

類似性の計測結果は、ユーザの音響的応答に最も近いフレーズか、あるいは何らかの一致を生じ損なったことの表示の選択である。音響的フィードバック表示および望ましくは更に視覚的フィードバック表示がユーザへ与えられ、一致した

フレーズを識別し、かつそれが正しいかあるいは正しくないかを表示する。ユーザ応答は、1 つの単語、幾つかの単語、唯 1 つまたは幾つかのフレーズが教育過

程において突合はされる文章または複数の文章を含み得る。表示された誤りをどのように克服するかの付加的な教育情報もまた、音響的／視覚的な方法で提供されることが望ましい。望ましくはヘッドセット14(図1)とディスプレイ16の一部をなすヘッドフォン22がこの目的のために用いられることが望ましい。

次に、本発明の一実施形態による「予期音響的応答基準データベースB」の生成中の図1のシステムの動作の全体的機能ブロック図である図3を参照する。同図において、種々の地理的範囲、種々の年齢および両方の性の配分を考慮した複数のネイティブ・スピーカにより発声されるフレーズを記録するためにマイクロフォン30が用いられる。

各話者は、複数の予め定めたフレーズを発音する。複数の予め定めたフレーズの各々ごとに、各話者がフレーズを正確に発音し、更にこのフレーズを不正確にその都度複数の予め定めた発音の誤りの異なるものを数回繰返す。記録は、「暫定音響的手本データベースA」に保持される。このデータベースは、強化された統計的な基礎を提供するため各話者に対して先に発音された各フレーズの複数の記録を含んでいる。

単語ベースの音声認識が行われる時、発声されたフレーズ・パラメータが抽出されて、「予期音響的応答基準データベースB」に既に格納されたフレーズ・パラメータと合成されて「予期音響的応答基準データベースB」を構成する。このデータベースは、複数の予期応答基準を含み、該予期応答基準は、満足し得る発音を有する第1の複数の予期応答基準と、該満足し得る発音を有する第1の予期応答基準の各々に対して各々が異なる発音エラーを有する第2の複数の予期応答基準とを含んでいる。

複数のM人の話者の各々により各フレーズがN回正確に記録されることが判る。このフレーズは、M人の話者の各々によりそれぞれ異なる発音誤りを含むし個の異なる形態で更にN回記録される。

次に、本発明の別の実施の形態による音響的基準ライブラリの生成中に図1のシステムの動作の全体的機能ブロック図である図4を参照する。同図において、「予期音響的応答基準データベースB」は、音素の言語記録を生じるため用いら

れるテキストおよび音素の言語ファイルを生成することによってコンピュータ生成される。この音素の言語記録は音素データベースCと共に用いられて、「予期音響的応答基準データベースB」を一緒に構成するフレーズ・テンプレートを生成する。

図4の実施の形態において、フレーズ・テンプレートは典型的に単語または単語の組合せではなく、むしろ音素、二音素および音素間の遷移の如き発声要素の特徴の組合せである。音素ベースの音声認識において、認識される発声の特徴は最良の一一致を見出すためにこれら組合せと比較される。

次に、図2の全体的機能ブロック図による言語の教育中のシステムの動作を示す全体的フローチャートと一緒に構成する図5Aおよび図5Bを参照する。フローチャートに示される表示された初期の準備がいったん完了し、かつ望ましくはデータベースEから聴かれる音声の種類が選択された後に、レッスンが選択され、ユーザは選択された音を発音する方法の説明が与えられる。選択された各音ごとに、基準音響的手本データベースEからとられた基準の音響的手本が、ユーザからの予期音響的応答を引出すためにユーザに対して再生される。

ユーザの応答は、記録され、開示が参考のため本文に援用される米国特許第5,487,671号に記載される如き「生徒の応答手本レコーダ」により、「予期音響的応答基準データベースB」に含まれる予期応答基準と比較される。

最良の一一致が正確な応答に対するならば、ユーザに対して正確なフィードバックが行われ、レッスンは次の音響的手本へ進む。

最良の一一致が基準が発音の誤りを含む予期応答基準に対するならば、適切なフィードバックがユーザに対して行われる。このフィードバックは、誤りの説明、それを正す方法と共に、予期応答基準の再生を含んでいることが望ましい。本発明の望ましい実施の形態によれば、誤って発音されたフレーズが「基準音響的手本データベースF」からユーザに対して再生される。

「ユーザ・フォローアップ・データベースD」は、ユーザの進歩を示すためまたは他の目的で最後即ち最も新しいユーザ応答を再現するために用いられる。これは、システム・フィードバック中に含まれる。

次に、図3の全体的機能ブロックによる言語の教育のための音響的基準ライブラリ生成中のシステムの動作を示す全体的フローチャートと一緒に構成する図6A、図6Bおよび図6Cを参照する。

フローチャートに示される初期の用意がいったん完了すると、訓練された話者が正確なフレーズと、発音における1つ以上の誤りがなければ発音が正確なフレーズに類似する複数の不正確なフレーズを発声して、それぞれが異なる発音誤りを含む予期応答基準を生じる。このような各組の正確なフレーズと不正確なフレーズが記録される。本発明の望ましい実施の形態によれば、暫定音響的手本データベースAが種々の記録を含む。データベースAは、図3に関して先に述べたように、単語ベースの音声認識に用いられる図6Cの「予期音響的応答基準データベースB」を生じるために用いられる。

次に、図4の全体的機能ブロックによる言語教育のための音響的基準ライブラリの生成中のシステムの動作を示す全体的フローチャートである図7を参照する。同図において、簡単なテキストと音素言語を入力して、テキストを表示された音素言語へ変換するためコンピュータが用いられる。先に述べた種類の音素データベースCを用いて、フレーズ・テンプレートが生成される。このフレーズ・テンプレートは、次に「予期音響的応答基準データベースB」に格納される。記述されたこのプロセスは、システムにより用いられる各フレーズ・テンプレートに対して実施される。フレーズ・テンプレートが典型的に単語または単語の組合せではなくむしろ音素、二音素および音素間の遷移の如き発声要素の特徴の組合せであることが判る。音素ベースの音声認識においては、認識される発声の特徴が最良の一一致を見出すために、これらの組合せと比較される。

次に、本発明の望ましい実施の形態による図4および図7において用いられた種類の「音素データベースC」の生成を示す図8および図9を参照する。典型的には、例えば図9に示された種類のラベルを付した発声のデータベース50は、

eメール・アドレス online-service@ldc.upenn.eduで米国ペンシルバニア州大学のLinguistic Data Consortiumから入手可能である、TIMIT Acoustic-Phone

tic Continuous Speech Corporaから得ることができる。典型的に、eメール・アドレス sales@entropic.com で Entropic Cambridge Research Laboratories 社から入手可能な HTK (Hidden Markov Model Toolkit) の如き市販のソフトウェアに具現されたテンプレート・ビルダ 52 は、データベース 50 上で動作し、「音素データベース C」を提供する。図 8 の技術は、種々の言語に適用可能である。

音素データベース 58 が種々の言語からの音素を含む場合は、「音素データベース C」は、図 10 に示されるように、複数の音素データベース 54、56 を組合わせることにより実現される。学習されあるいは発声される言語ならびにユーザの生来の言語の音素を含む音素データベース 54、56 がこのように組合わされて強化された音声認識を提供することが本発明の特定の特徴である。

次に、音素を用いる音声認識の図示である図 11 を参照する。図示された事例において、予期される単語は「t o m a t o (とまと)」である。予期される代替的な発音のネット (net) が生成される。この場合、話者は最初の「o」を「O」、「OW」、あるいは「U」と発音し得、発音「O」が正しいと見なされる。

同様に、ユーザは「a」を「A」または「E Y」と発音し得、発音「E Y」が正しいと見なされる。

図 11 は、全ての音素が单一の言語に帰属する音声認識に対して用いられることが特徴付けられる。

次に、種々の言語の音素を用いる音声認識の図示である図 12 を参照する。当例は、日本語を母国語とする話者によって発声された英語を認識するように設計されている。この場合、予期される単語は「L o s A n g e l e s」における如き「L o s」である。この場合、話者が「L」を「L」（丸で囲んだ「L」）

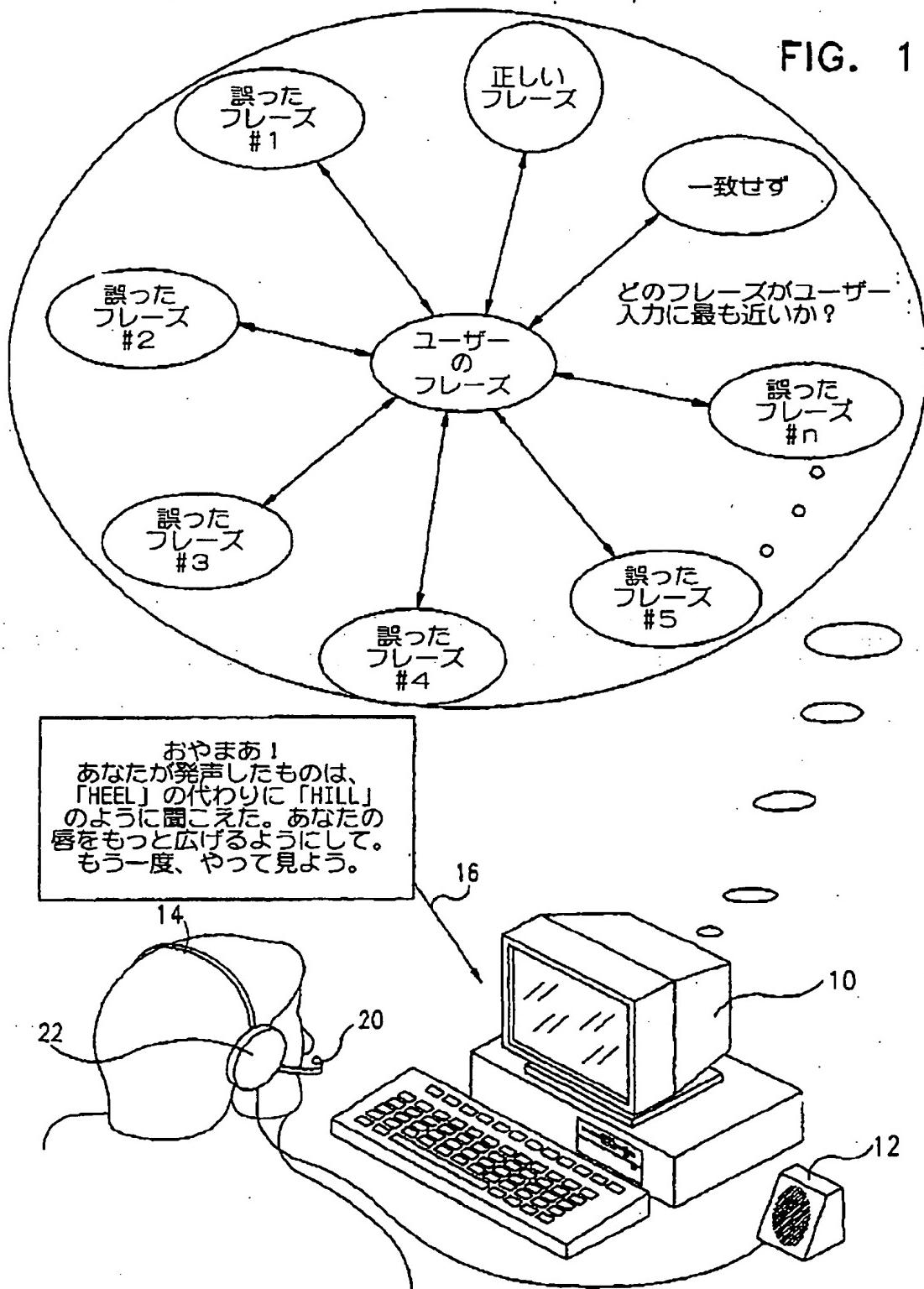
英語の「R」（丸で囲んだ「R」）、あるいは日本語の「R」（四角で囲んだ「R」）と発音し得ることが判る。

図12は、音声認識のため用いられる全ての音素が单一の言語に帰属しないことが特徴付けられる。図12の事例では、音素の一部が英語の音素（丸で囲んだ文字）であり、音素の一部が日本語の音素（四角で囲んだ文字）である。

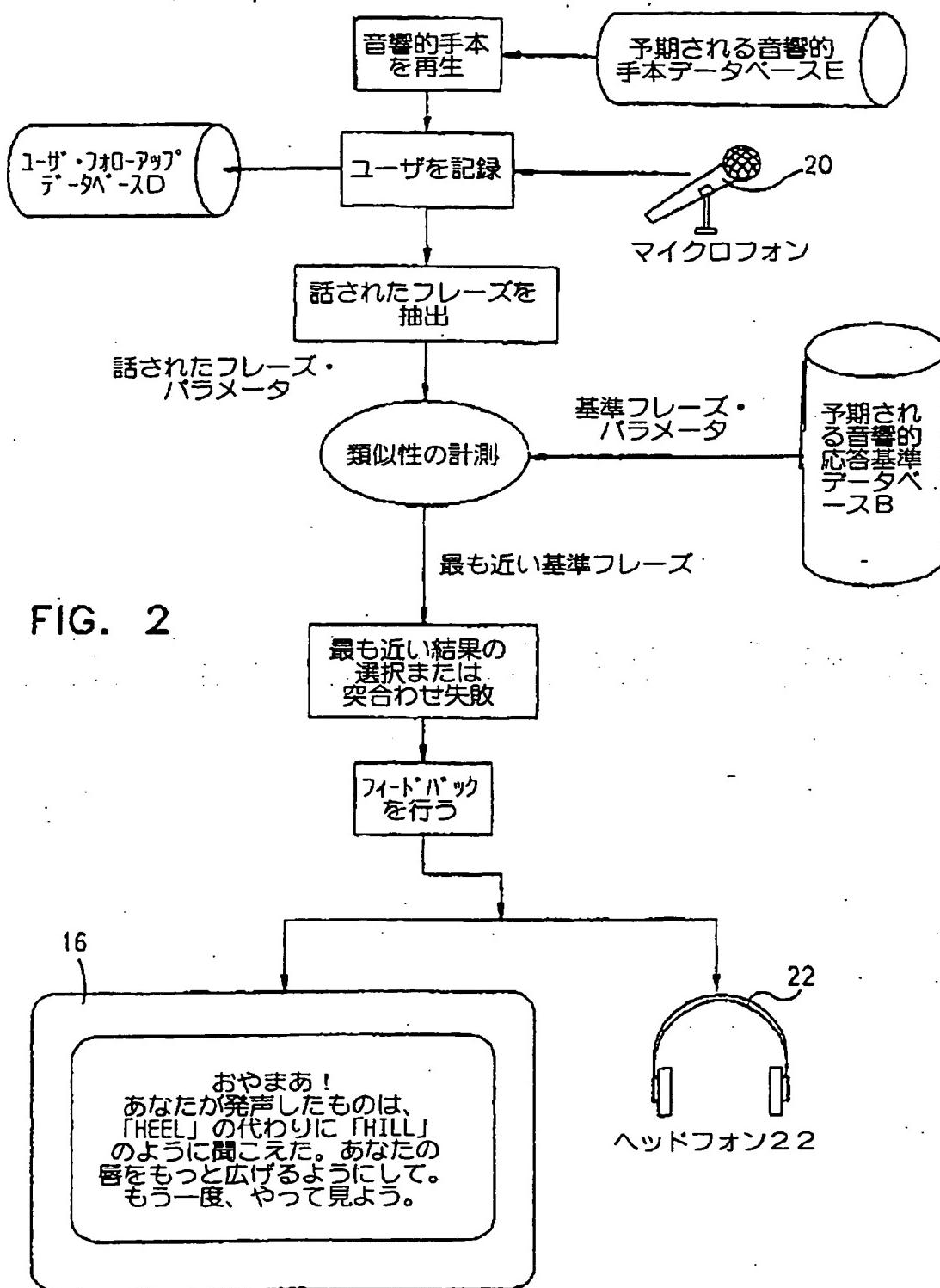
このように、言語の教育のため図12の音声認識技術を用いる時、生来の日本語の特徴的な誤った発音がシステムにより認識され、必要な教育フィードバックがユーザに対して行われることが判る。図12の音声認識技術が他の音声認識用途に用いられる時、この技術は英語の発音が認識されるのに完全ではない日本語を母国語とする日本語話者により英語が発声されることを可能にする。

当業者には、本発明が本文に特に示され記述されたことに限定されないことが判るであろう。むしろ、本発明の範囲が本文に述べた種々の特徴および要素の組合せおよび部分的な組合せの両方、ならびにその変更および拡張を含むものである。

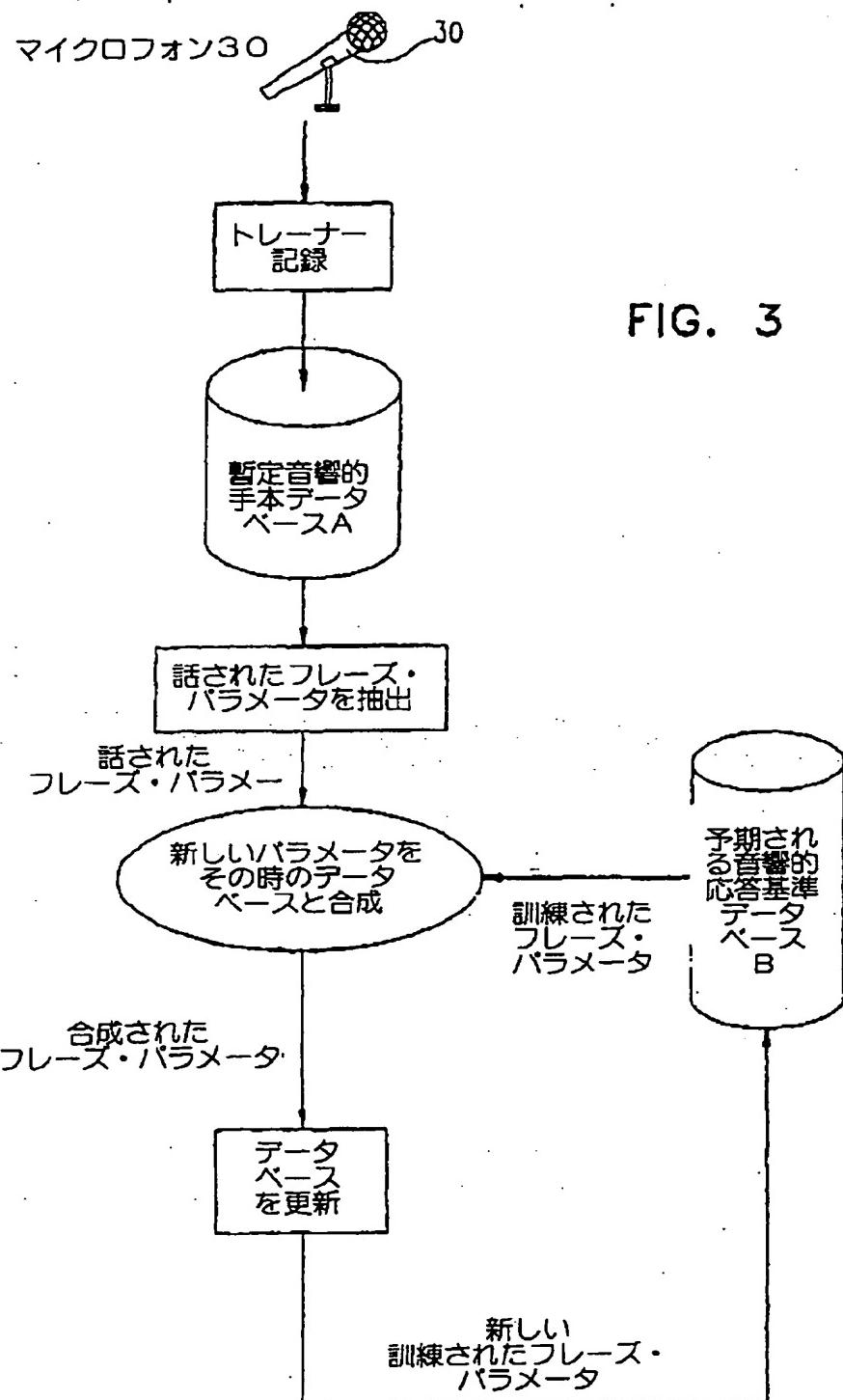
【図1】



【図2】



【図3】



【図4】

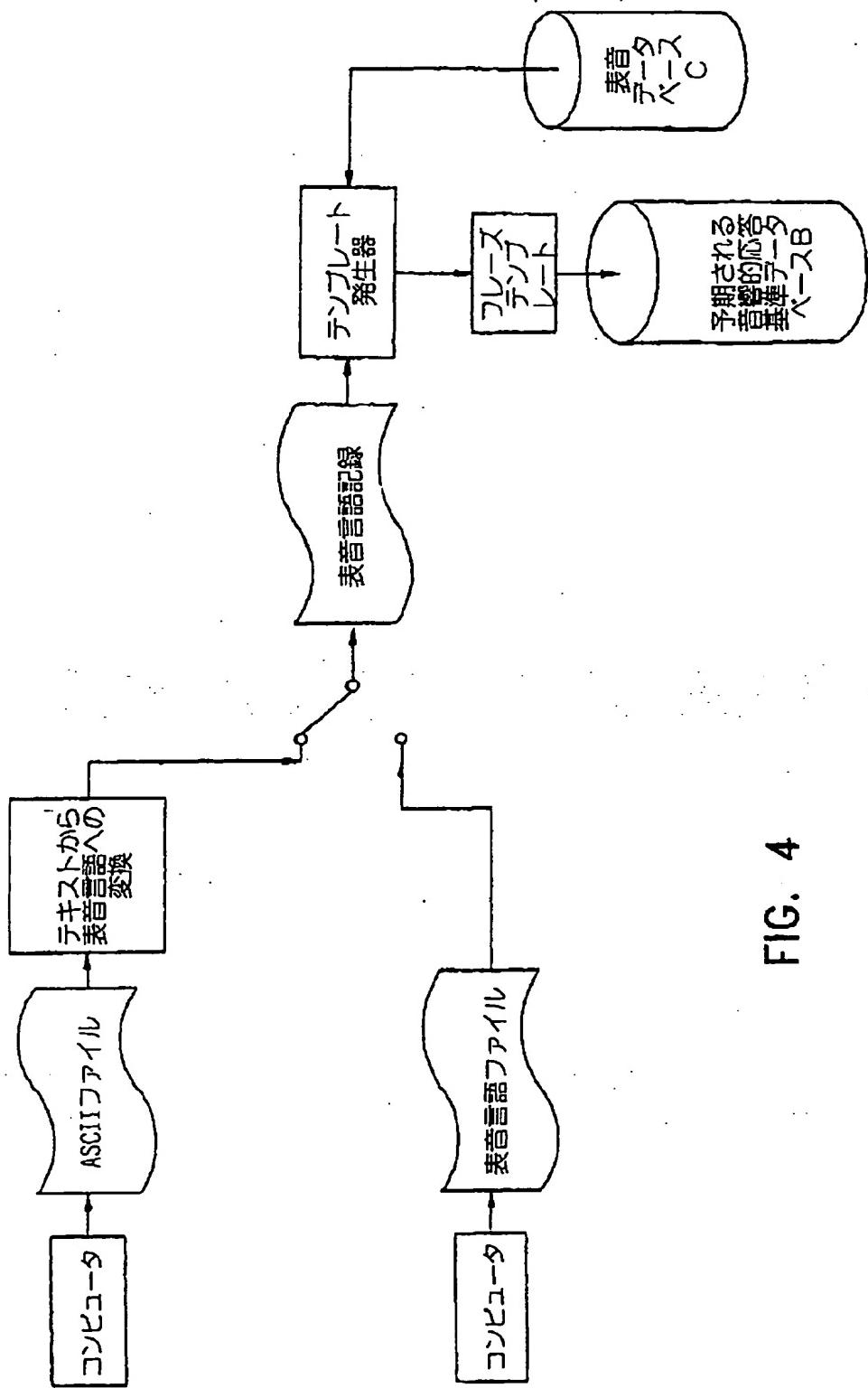


FIG. 4

【図5】

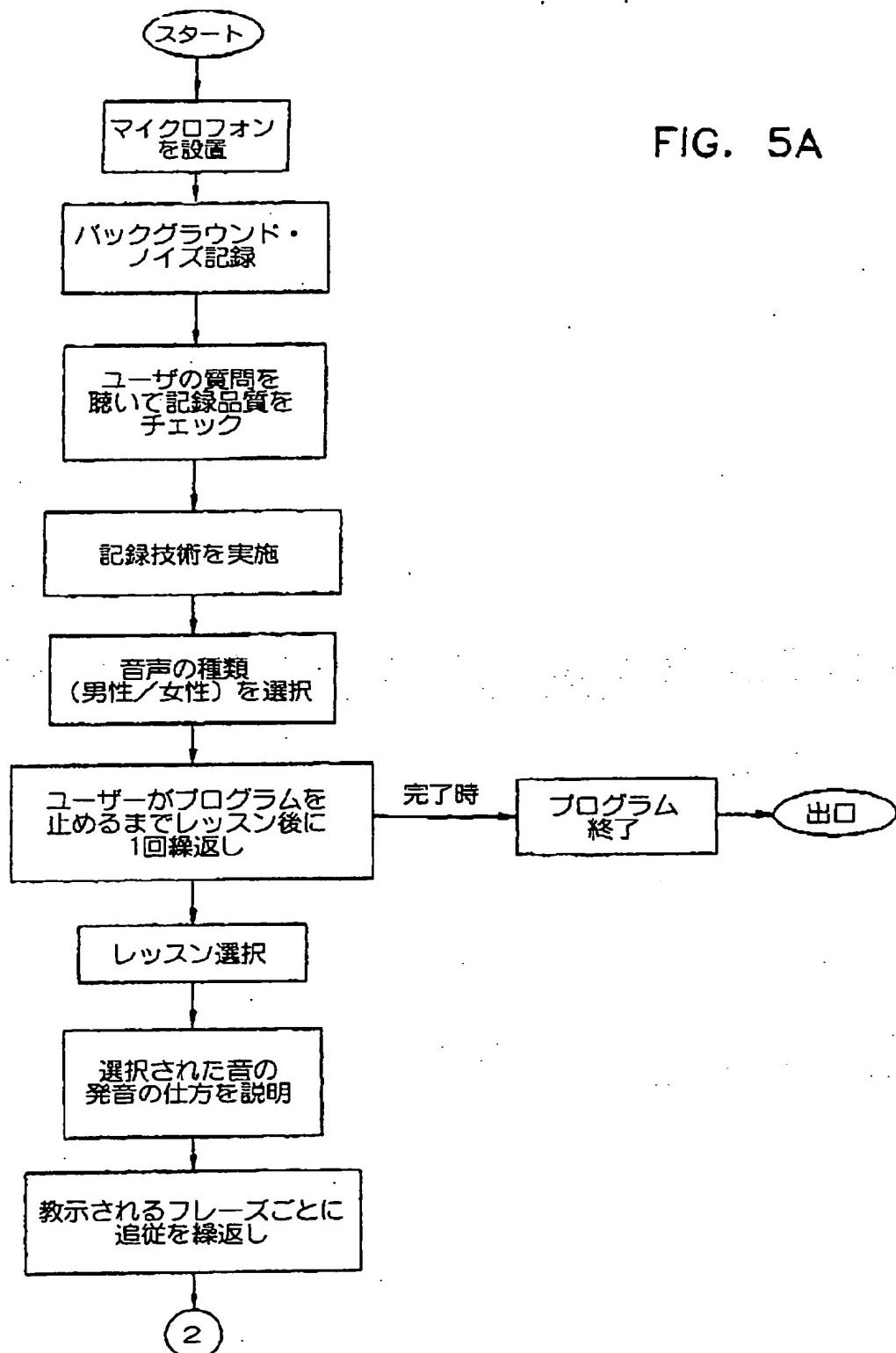
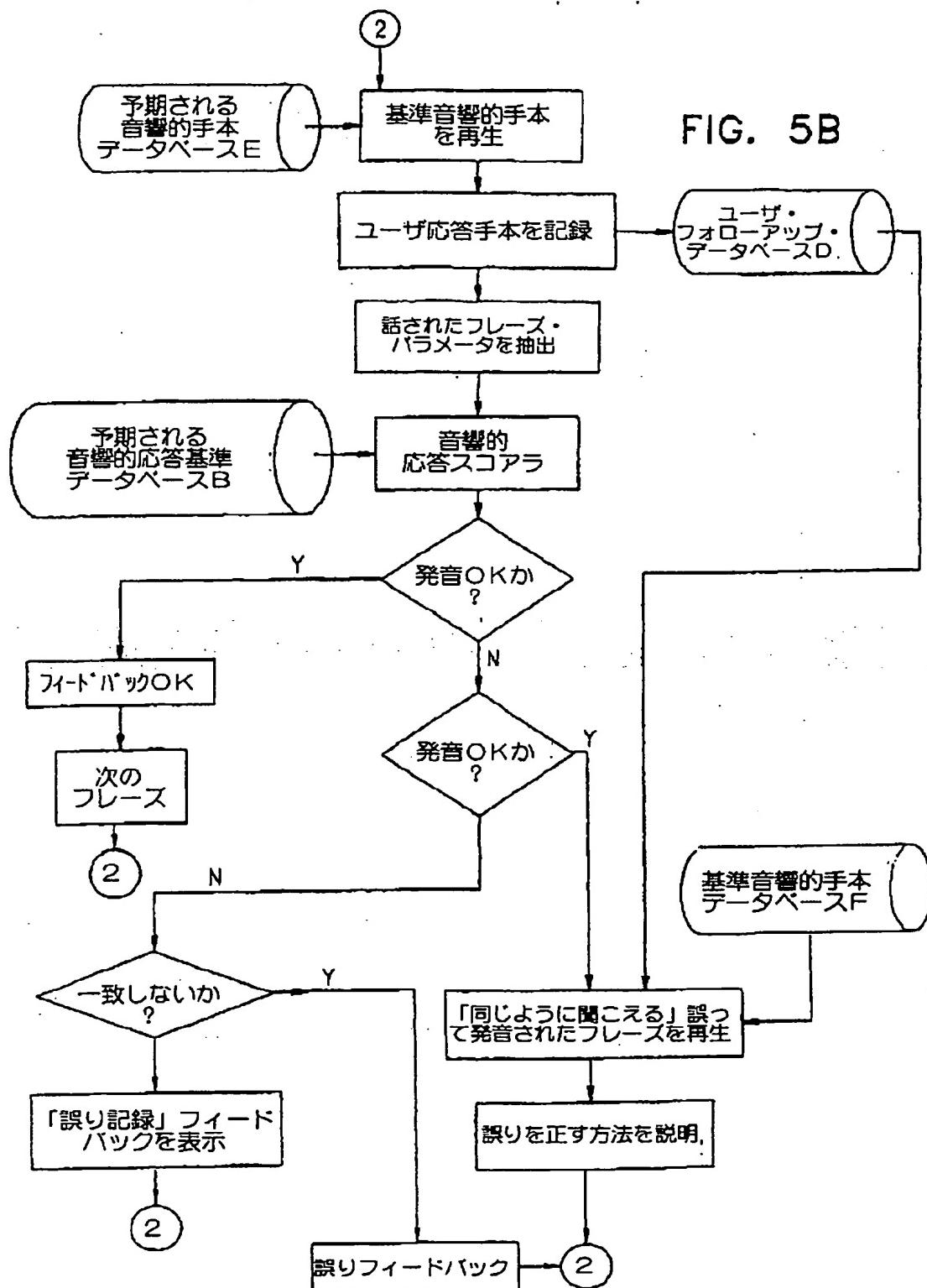


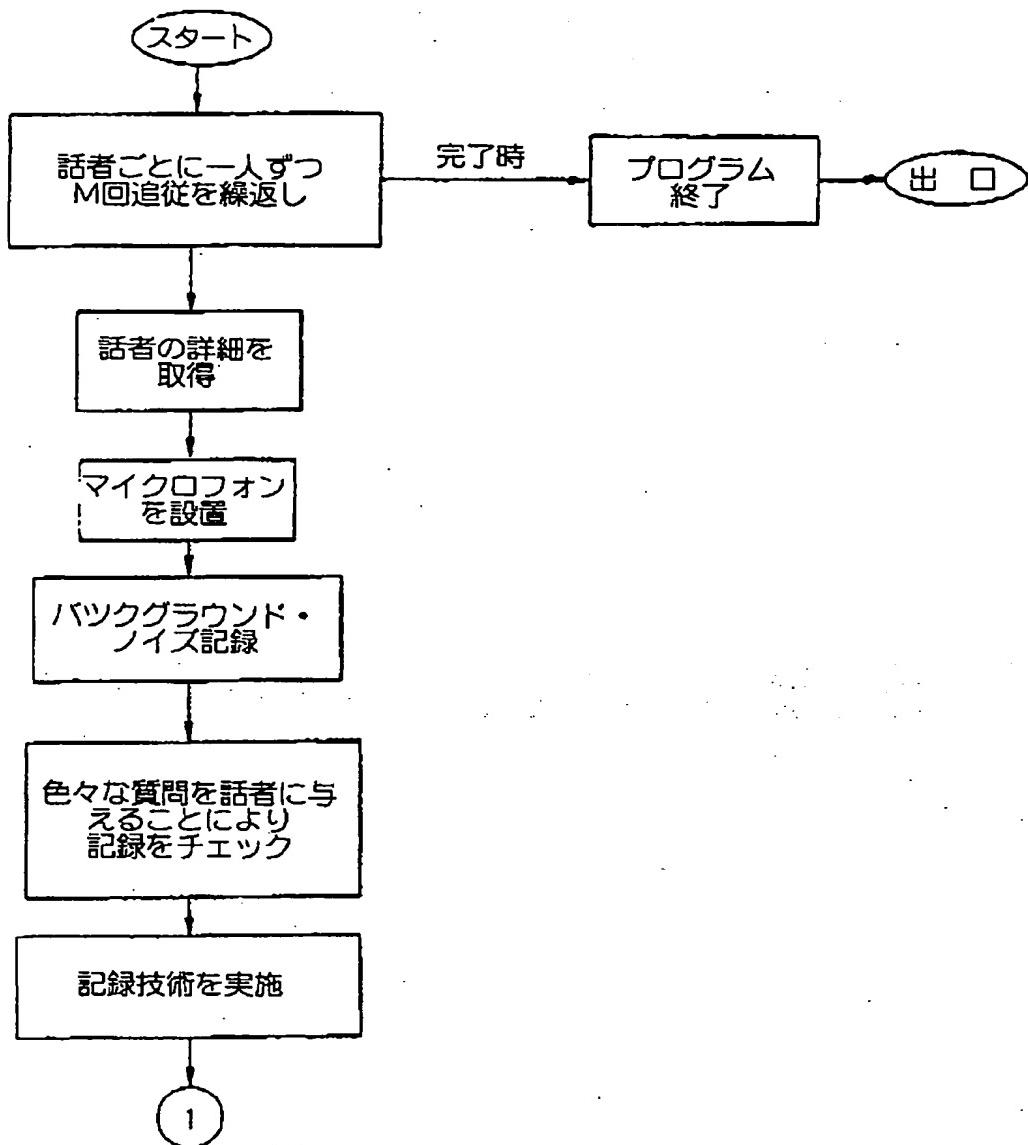
FIG. 5A

【図5】

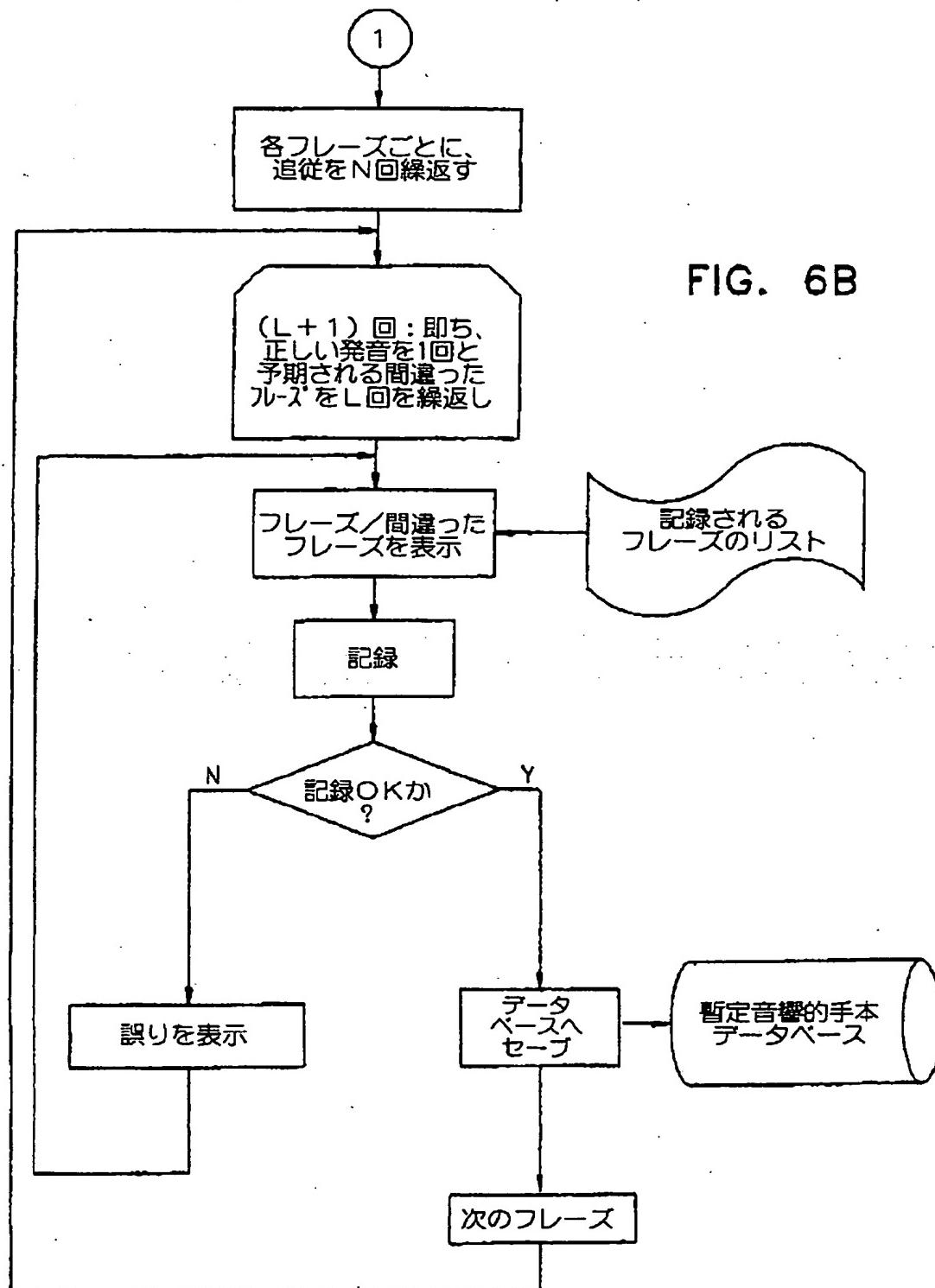


【図6】

FIG. 6A

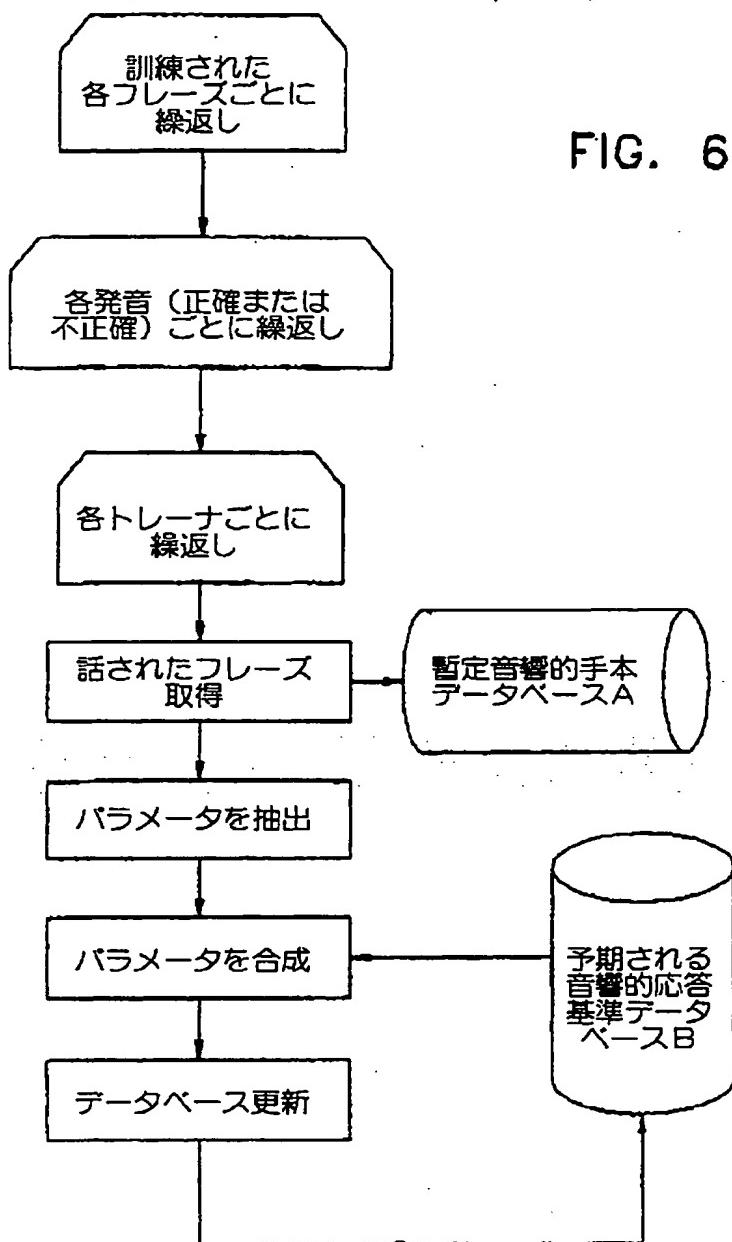


【図6】



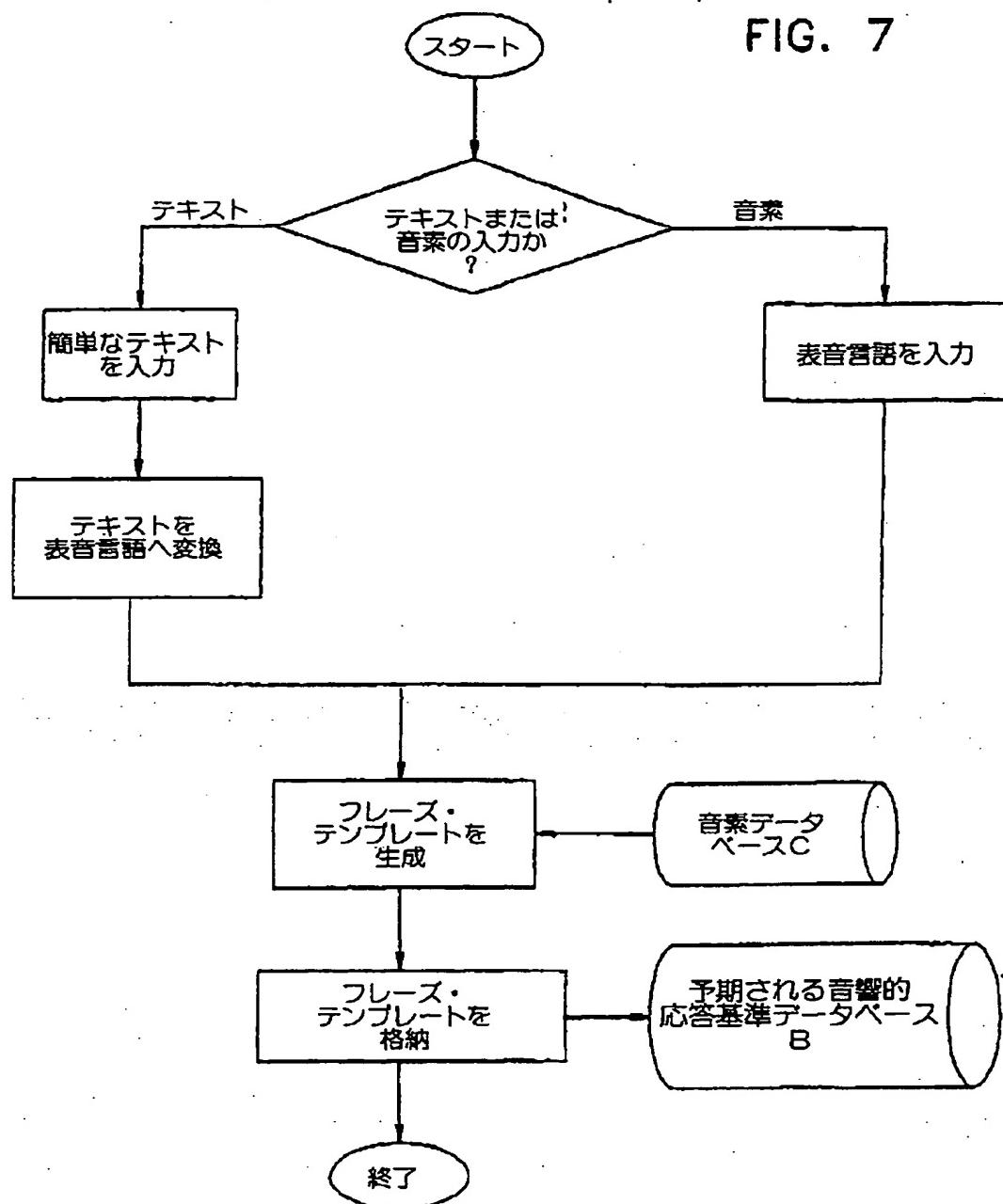
【図6】

FIG. 6C



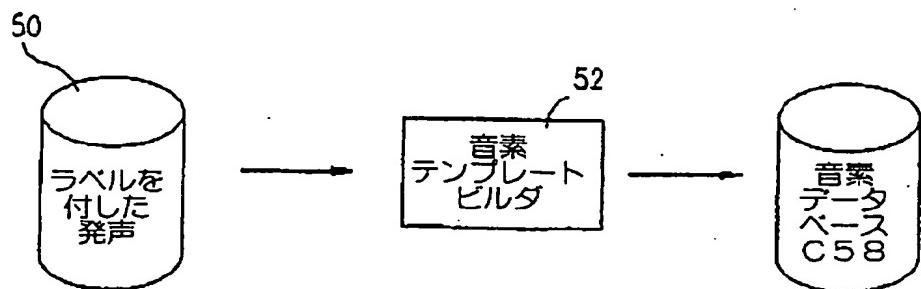
【図7】

FIG. 7



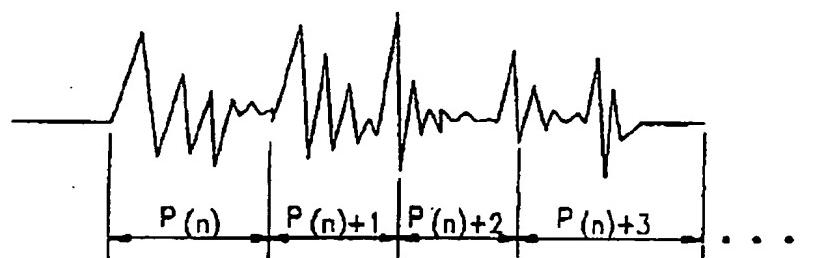
【図8】

FIG. 8



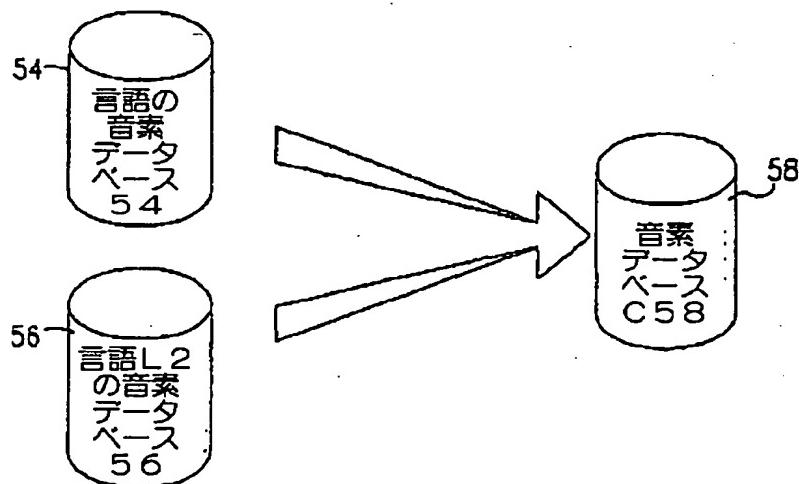
【図9】

FIG. 9



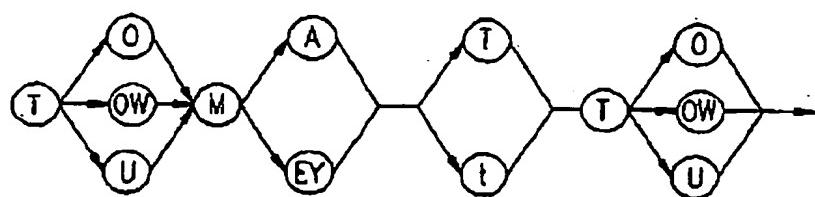
【図10】

FIG. 10



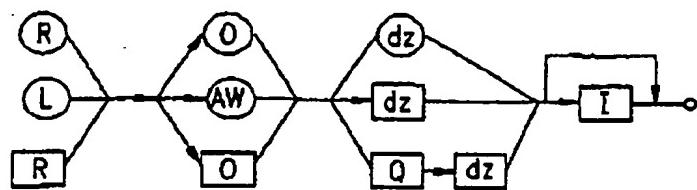
【図11】

FIG. 11



【図12】

FIG. 12



## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/IL97/00143

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G09B 1/00, 5/00, 19/00, 04, 06  
US CL : 434/156, 157, 167, 169, 185

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 434/156, 157, 167, 169, 185

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

GPI Web Client

Search Terms: foreign, voice, recognition, words, correct, incorrect, database

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claims No.
X	US 5,487,671 A (SHPIRO et al) 30 January 1996, whole document	23-31
&	US 5,503,560 A (STENTIFORD) 2 April 1996.	1-31
&	US 5,393,236 A (BLACKMER et al) 28 February 1995.	1-31

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not to conflict with the application but cited to understand the principle or theory underlying the invention
"A"		document defining the general state of the art which is not considered to be of particular relevance
"E"	X	earlier document published on or after the international filing date
"L"		document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified)
"O"		document referring to an oral disclosure, use, exhibition or other means
"P"		document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search  24 OCTOBER 1997	Date of mailing of the international search report  12 NOV 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOHN ROVNAK Telephone No. (703) 308-3087

Form PCT/ISA/210 (second sheet), July 1992\*

---

フロントページの続き

(81) 指定国      E P (AT, BE, CH, DE,  
DK, ES, FI, FR, GB, GR, IE, IT, L  
U, MC, NL, PT, SE), OA (BF, BJ, CF  
, CG, CI, CM, GA, GN, ML, MR, NE,  
SN, TD, TG), AP (GH, KE, LS, MW, S  
D, SZ, UG), UA (AM, AZ, BY, KG, KZ  
, MD, RU, TJ, TM), AL, AM, AT, AT  
, AU, AZ, BA, BB, BG, BR, BY, CA,  
CH, CN, CU, CZ, CZ, DE, DE, DK, D  
K, EE, EE, ES, FI, FI, GB, GE, GH  
, HU, IL, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, M  
D, MG, MK, MN, MW, MX, NO, NZ, PL  
, PT, RO, RU, SD, SE, SG, SI, SK,  
SK, TJ, TM, TR, TT, UA, UG, US, U  
Z, VN, YU